

# Ontology for Semantic Integration in a Cognitive Surveillance System

FULL PAPER — Anonymous SAMT 2007 submission

**Abstract.** The recent growing of interest in sophisticated intelligent systems dealing with advanced interfaces, as in the case of Cognitive Vision Systems (CVS), motivates the apparition of ad-hoc stages designed for the integration of multiple domains of semantic knowledge. This paper proposes a novel ontology as a suitable mechanism to restrict and integrate high-level semantics for Human Sequence Evaluation (HSE), which incorporates multilingual capabilities and multipurpose end-user interfaces. The main contributions of this paper are the conception of a neutral semantic layer formed by High-Level Semantic Predicates (HLSP), which allow to link semantics representations from vision and linguistic domains; and the use of *situations* instead of verbs as the basic elements for an ontological categorization of interpreted occurrences. In this particular approach, the domain has been restricted to the field of human behaviors in outdoor surveilled scenarios, involving interactions among pedestrians, static objects, and vehicular traffic.

## 1 Introduction

The use of the expression *Multimedia Surveillance System* (MSS) is becoming extended when referring to the new generation of applications which deal with different media streams and aim to automatically retrieve content from a controlled environment, by means of the integration of video technology and sensor networks [7]. Since the different continuous media streams from real-time monitoring generate abundant low-level data, there exists an increasing interest in finding methods for the extraction of meaningful content, so that the educed semantic knowledge can be forwarded to high-level reasoning stages. A question arises next about how to properly link these stages to advanced end-user interfaces for the use of semantic content.

There exists a growing involvement of the research community into Cognitive Vision Systems (CVS) engineering, which operate on different levels of abstraction and analysis related to human cognition. These systems enable to exploit current machine capabilities in two different new manners: first, to achieve *interpretations* of the symbolic descriptions which are closer to human perception; and second, to implement *advanced user interfaces* aiming to enhance communication with and control from a final user, e.g. by means of visual representations, Natural Language (NL) text, and speech interaction.

This is well accomplished by the general conception of *Human Sequence Evaluation* (HSE) proposed by González [9], in which the interpretation of human

behaviors in image sequences is performed by a modular architecture of a CVS, thus allowing to develop multiple user-oriented applications, in particular MSS. In such a framework, the development of proper criteria for high-level knowledge sharing and validation have been seen as essential, in order to achieve an efficient handling of these many different tasks [4]. Due to the broad spectrum of semantic representations that have come to light in the last years, it is necessary to find mechanisms that clarify the structure of knowledge in particular domains, and to devise effective knowledge representations towards integration purposes. Towards this end, ontologies have been widely accepted as convenient tools.

This contribution addresses the use of ontologies as an integrative framework for knowledge representation, within a HSE system with multiple user interfaces and multilingual capabilities. We also discuss several criteria to model the semantic background of such ontologies. Ontologies are thus used to link the different semantic representations at cognitive, high-level stages.

Next section reviews some of the recent work done in human behavior interpretation, involving high-level symbolic and semantic stages and related to the field of Cognitive Vision. Section 3 briefly introduces the HSE system and gives a short description of the different types of semantic representation that are implied. Section 4 discusses some considerations for the use of ontologies in cognitive environments, and proposes taxonomies for the universe of semantic entities and situations restricted to the domain of interest. Section 5 concludes the paper and points out some improvements to be done in the future.

## 2 Related Work

The evaluation of human behaviors in image sequences is a commonly required task for applications such as surveillance, content-based retrieval of documents, or advanced interfaces related to cognitive fields. Nevertheless, while low-level visual techniques have been actively investigated for decades, high-level processing has acquired significant attention in the field of vision specially in the last years, as stated in [17].

Current CVS build on both purposive and reactive data flows, which incorporate techniques from several vision and reasoning levels. Mechanisms for the evaluation, gathering, integration and active selection of these techniques are fundamental to attain robust interpretation of dynamic information [22, 10]. Nagel has actively investigated the field of CVS applied to vehicular traffic surveillance [2, 1]. He tackles the high-level analysis of visual occurrences by means of fuzzy logic inference engines, and derives the results to the generation of NL textual descriptions. In [3], Buxton reviews progress in generative models for advanced CVS to explain activities in dynamic scenes, observing applications such as education, smart rooms, and also surveillance systems. The study of interactions among moving people and objects is undertaken under statistical approaches for high-level attention and control. In [16], Maillot et al. argue against application-specific designs of knowledge bases for vision systems, and

suggest a general methodology to hide the low-level complexity on CVS by using visual concept ontologies.

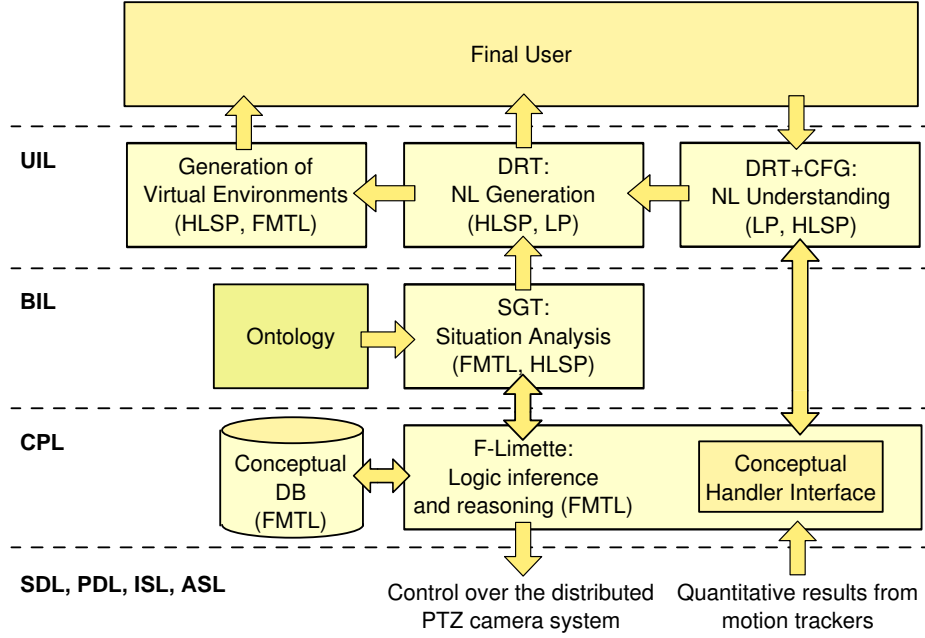
Crowley proposes some conceptual frameworks towards the understanding of observed human activity, including interactions [6]. He suggests a model consisting of a set of roles to be accomplished by entities; specific configurations of interrelated entities playing roles conform the so-called situations. The EU Project ActIPret uses semantic-driven techniques to automatically describe and record activities of people handling tools in NL, by exploiting contextual information towards symbolic interpretation of spatiotemporal data [22]. Its reasoning engine focuses on the coordination of visual processes to obtain generic primitives from contextual control. In [17], Park and Aggarwal discuss a method to represent two-person interactions at a semantic level, also involving user-friendly NL descriptions. Human interactions are represented in terms of cause-effect (event) semantics between syntactical agent–motion–target triplets. The final mapping into verb phrases is based on simultaneous and sequential recognitions of predefined interactions. The intelligent multimedia storytelling system CONFUCIUS interprets NL inputs and automatically generates 3D animation and speech [15]. Several methods for categorizing eventive verbs are discussed, and the notion of visual valency is introduced as a semantic modeling tool.

This need for coordination of contextual knowledge suggests to single out specific stages for semantic manipulation. Although many advanced surveillance systems have adopted semantic-based approaches to face high-level issues related to abstraction and reasoning, the use of ontologies at high levels of such systems is not yet commonly implemented.

### 3 Cognitive Surveillance System for HSE

An HSE system aims to extract automatic descriptions of human behavior from image sequences in restricted discourse domains, in this case urban outdoor surveillance environments. The final goal is the evaluation of complex behaviors involving humans, vehicles, and static objects.

Another target entailed by HSE is to facilitate the presentation and exchange of information with final users, by means of multipurpose interfaces. In our particular case, this is achieved by providing three different interfaces: (i) automatic generation of descriptions of the interpreted occurrences in multiple natural languages; (ii) NL query retrieval interface to gain feedback from the final user, also accepting multiple natural languages for the inputs; and (iii) automatic generation of animations which show virtual representations of the behaviors and occurrences involved in the scene. Fig. 1 depicts the general architecture of the system, and Fig. 2 shows a captured frame that gives an example of surveilled scenario.



**Fig. 1.** General HSE architecture of the CVS for surveillance application considered. The knowledge representation formalism used by each module is enclosed in parenthesis. The acronyms on the left side stand for the different levels implied, from bottom to top: Active Sensor (ASL), Image Signal (ISL), Picture Domain (PDL), Scene Domain (SDL), Conceptual Primitives (CPL), Behavior Interpretation (BIL), and User Interaction (UIL) levels.

### 3.1 Domains of Semantic Representation

Several formalisms are employed by a HSE system in order to represent semantic knowledge. These representations are conditioned to the application domain they address. Given the general architecture of the system in Fig. 1, we next discuss the semantic representations implied. These are Fuzzy Metric-Temporal Horn Logic (FMTL), High-Level Semantic Predicates (HLSP), and Linguistic Predicates (LP).

**Fuzzy Metric-Temporal Horn Logic (FMTL)** facilitates a schematic representation of conceptual knowledge which is time-delimited and incorporates uncertainty. It is found at the Conceptual Primitives Level (CPL), see Fig. 1. This formalism is used by an inference engine for two main goals: to provide a geometrical description of the scene (prior knowledge), and as a basis to infer higher abstractions upon the limited facts provided by the Computer Vision levels (trajectories and other temporal-restricted information). In order to deduce



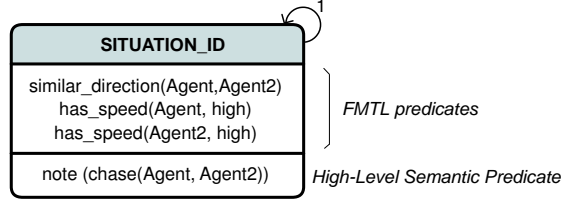
**Fig. 2.** Snapshot extracted from the recordings on one of the possible scenarios covered by the system. In this scene, a pedestrian is stealing an object from another pedestrian.

new knowledge, the inference engine requires the presence of conceptual models to describe the domain of interest. A conceptual database of knowledge is being updated continuously, since the retrieved facts from motion evaluation vary at every time-step. See [19, 1] for extended information.

In the current implementation, FMTL is manipulated by the inference engine *F-Limette* [19] to represent and reason about spatiotemporal developments. Uncertainty is treated by assigning fuzzy degrees-of-validity to the quantitative values generated by the motion trackers [18]. Next example shows a metric-temporal modeling for the inference of a new FMTL predicate upon the quantitative values for the orientations of two agents.

```
always(similar_direction(Agent, Agent2):-
    has_status(Agent,_,_,_,Or1,_),
    has_status(Agent2,_,_,_,Or2,_),
    Dif1 is Or1 - Or2,
    Dif2 is Or2 - Or1,
    maximum(Dif1, Dif2, MaxDif),
    MaxDif < 30
).
```

**High-Level Semantic Predicates (HLSP)** gather those general formulae in First-Order Logic formalism which express semantic relations among the individuals, at a higher level than metric-temporal basic relations. They are used at the Behavioral Integration Level (BIL), see Fig. 1. HLSP are the result of a post-processing stage over conceptual facts, in which situational and behavioral models are applied. These new constraints are not only based on factual predicates and basic relations, as FMTL predicates do, but they also embed



**Fig. 3.** Situation scheme from a SGT. Once a set of FMTL predicates is asserted within a contextualized situation, a HLSP is generated. One must note that the situation scheme is available for evaluation only when the behavioral model of the SGT enables it, by means of specialization and prediction edges, see [1] for details.

situational restrictions which take into account *contextualization*, *integration*, and *interpretation* tasks. Hence, the set of HLSP reaches the higher account of semantics, in the cognitive sense that each one of them implies a perceived situation or behavior which is meaningful and remarkable by itself in the selected domain.

The current implementation for the generation of these high-level predicates is based on Situation Graph Trees (SGTs), see [1]. The nodes of these graphs are situation schemes which embed the contextual state of an agent at a discrete point of time, by relating a set of necessary FMTL facts to the situation. When the entire set of facts defined in the situation is asserted, a new interpretation for the scene is generated in form of a HLSP. SGTs are traversed at every time-step, and therefore the produced interpretations in HLSP are subjected to temporal validity.

**Linguistic Predicates (LP)** are conceptual representations of a linguistic-oriented semantic knowledge. In our particular case, LP are incorporated into Discourse Representation Structures (DRS), which are part of the Discourse Representation Theory [12]. These structures provide a suitable representation formalism to cover the gap between the conceptual and linguistic stages of knowledge. They are used for NL generation and understanding at the User Interaction Level (UIL), see Fig. 1. They offer a logic-based linguistically oriented representation of the temporal development within a scene, which is eventually transformed into NL texts [8, 1]. Nevertheless, the problem of converting conceptual predicates into NL has not yet been solved in a generalized way.

DRS are constituted by a set of referents and a universe of conditions. In our case, the referents are chosen from the set of instantiable entities which will be defined by the ontology, and the conditions are conformed by a subset of predicates linguistically oriented towards a particular language. Each of these *linguistic predicates* requires distinct thematic arguments depending on the particular language and situation. Linguistic predicates from different languages describing a single situation are related to a single HLSP; the ones describing

	FMTL	HLS	LP
Type of semantics	Metric-temporal (basic relations)	Thematic roles (inferential role semantics)	Linguistic-oriented (NL semantics)
Implementation	F-Limette	Situation Graph Trees (SGTs)	Discourse Representation Structures (DRSs)
Models implied	<i>Prior knowledge:</i> scenario models <i>Inference:</i> human motion models	Behavioral models (contextual and intentional)	Linguistic models (Syntax, morphology, alignment, etc.)
Benefits	Allows inference of higher-level predicates upon asserted facts	Linguistic-oriented, highest level of interpretation	Facilitates to convert between logic and NL
Limitations	Limited to metric- temporal reasoning	Domain-dependent and target-oriented	Language-dependent

**Table 1.** Table of semantic representations in HSE [9].

entities are identified with those entities. Examples of DRS including linguistic predicates in English and Spanish are shown next<sup>1</sup>.

$$e_1 : \langle \{x, n, t_1, e_1\}, \{theft(x), e_1 : detect(x), t_1 < n, e_1 \subseteq t_1\} \rangle$$

NL: “A theft was detected”

$$\langle \{x, y\}, \{peatón(x), peatón(y), perseguir(x, y)\} \rangle$$

NL: “Un peatón persigue a otro peatón”

Table 1 contains a summary of some remarkable features for the different semantic representation formalisms described. After analyzing the properties of the different types of knowledge, we have chosen to focus on HLSP for the construction of the ontology, since they contain the higher level of semantic information for our purpose and are independent from specific languages. In addition, they are a natural intermediate stage between metric-temporal logic from vision and NL semantics from linguistics, and thus suitable for a neutral representational framework.

## 4 Ontologies for integration of knowledge

An ontology is an explicit specification of a conceptualization [11]. Ontologies give a formal description of (i) the concepts that are implied in a domain, and

<sup>1</sup> In the first example, the condition between temporal referents  $t_1 < n$  characterizes the past tense for the NL generation.

(ii) the relationships that hold among these concepts. The main motivation for the use of ontologies is to *capture the knowledge involved in a certain domain of interest*, by specifying some conventions about the content implied by this domain [20]. Ontologies are especially used in environments requiring to share, reuse, or interchange specific knowledge among entities involved in different levels of manipulation of the information.

In the context of this paper, ontologies hold several interesting properties which make them adequate for the requirements exposed. In the first place, they establish agreements about knowledge, such as explicit models and assumptions of the application domain. As a consequence, a reduction of variability can be achieved in symbolic manipulation and abstraction processes, since a set of constraints is established for these frameworks: the *validity* of the models is thus defined. Secondly, these specifications are independent from the chosen tools, since they are qualified in conceptual terms, and hence do not affect the concrete implementation. In addition, this particularity avoids the possibility of redundant knowledge. Finally, due to their high hierarchization, ontologies permit to easily extend and modify their contents for refinement.

In this paper, we aim to design the ontology in a central framework of *cognitive semantics*, which is generally known by three tenets [5]:

- Language draws upon general cognitive resources, not upon a special or independent module.
- The conceptual structure of semantics is motivated by its usage.
- Semantics is not purely truth-conditional; in Langacker’s words, semantics means conceptualization [14].

As a consequence, we consider that the design of a proper ontology based on a cognitive background (i) may incorporate assumptions from cognitive linguistics, since the representation of this knowledge does not differ substantially from other conceptual structures; (ii) has to incorporate pragmatic and intentional models to approach to the speaker’s meaning; and (iii) the conceptual interpretations of facts resulting from these models shall be the basic elements of the ontological classification.

#### 4.1 Considerations for the Semantic Entities

CVS incorporate approaches for semantic representation of human behaviors. However, existing concept hierarchies are uniquely based in eventive verbs. As stated in [13], NL inherently incorporates concepts for actions, events, and states in form of verbs, and choosing an appropriate verb for an observed event can convey the meanings of the event effectively. Several efforts have been done for characterizing verbal semantic properties, especially for the English language. Nevertheless, the use of verbs as semantic linkers between languages does not seem to fulfill the requirements of multilingual applications. In some cases, verbs



cannot be directly translated into another language, since the closest source and target utterances have different senses or connotations. In some other cases, the description of a given situation is accomplished by using distinct natural language expressions which hold severe syntactical dissimilarities. The possible use of semantically close expressions requiring different arguments for the same situation, in one or more languages, makes difficult to completely adapt to a model for ontological categorization.

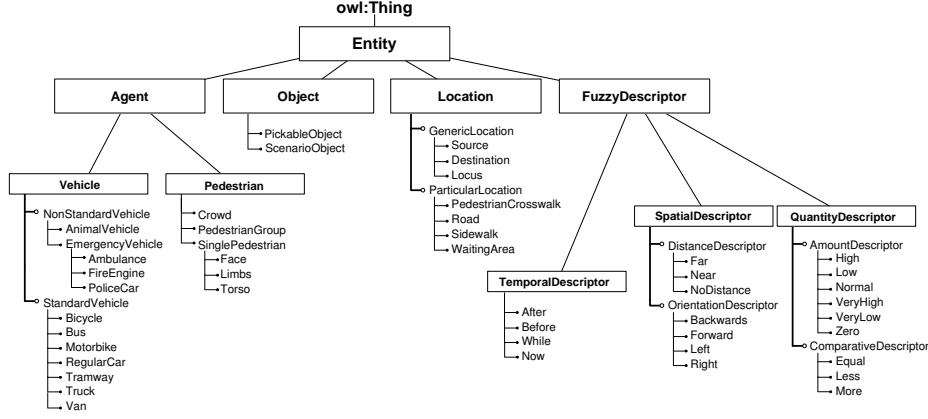
As a consequence, we have chosen to select *situations* instead of verbs as central elements of the ontology. Although many of these situations are expressed by verbs, e.g. *chase*, *yield*, or *escape*, there are many other interpretations of scenes which lead to complex mental representations but do not identify with any specific verb. This is the case of *DangerOfRunover*, a situation in which a person can suspect a dangerous behavior in development, or *AbandonedObject*, in which after a certain time of inactivity, a dropped object is considered as abandoned, in this case not regarding the identity of the person responsible of the act. These situations involve not only metric-temporal information and causation, but also behavioral models and attentional factors.

There exist many approaches for the ontological categorization of visually perceived events. An extensive review of the most important ones is done in [15], from which we remark Fillmore’s Case Grammar, Jackendoff’s Lexical Conceptual Structures, Dowty’s consideration of Thematic Proto-Roles, Fellbaum’s WordNet, Aspectual Classes proposed by Vendler, and Levin’s suggestion for verb classes. As an extension, our approach presented next relates each situation from the ontology with a set of required entities, which are classified depending on the thematic role they develop.

A taxonomy has been developed for the possible set of semantic entities in the described domain. The chosen list of entities include *agents* as those which can spontaneously act to change a situation, here pedestrians and vehicles; *objects* as static elements of the scene; *locations*; and a set of abstract *descriptors* which permit to add fuzzy modifiers to the conditions related to the entities. Other roles such as experiencer, goal, location, or instrument are easily enclosed in the selected categories. The described concepts of agent and object agree with the proposal by Crowley, namely *actor* and *prop* [6]. This taxonomy of entities, which will be used for the ontological classification of situations, is showed in Fig. 2. The main advantage of this approach is an independency of the particularities of verbs to a concrete natural language, thus facilitating the addition of multiple languages in the HSE system. It also enables the possibility to represent complex situations without restricting to the structure of a given language.

## 4.2 Ontological Categorization of Situations

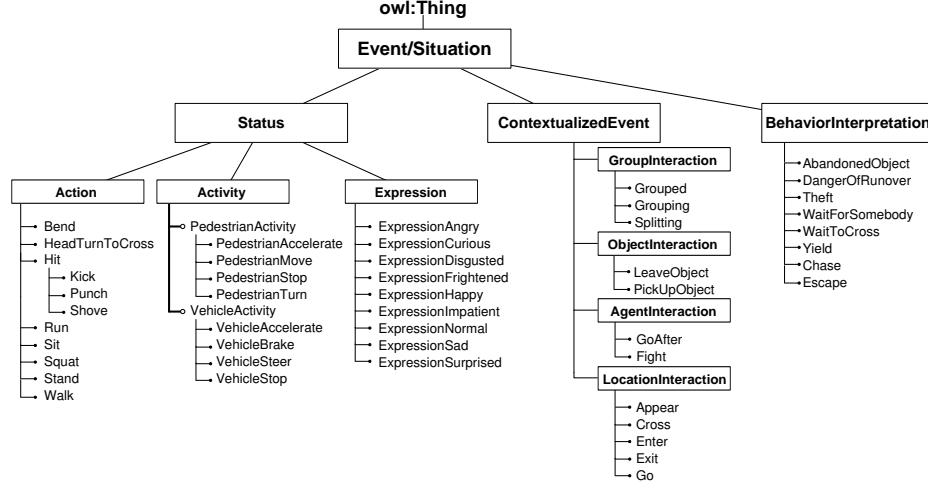
The main target for the proposed ontology is to enumerate and correlate the instantiable situations which are detectable in the selected domain, using a proper cognitive-based semantic representation. Now that the possible semantic participants have been established and organized, the set of situations can be classified.



**Table 2.** Example of taxonomy for the classification of Entities which are related to the targeted domain, see Section 4.1 for details. This taxonomy is part of the proposed ontology.

Talmy discussed the organization of conceptual material from a cognitive perspective in [21]. He tackled the problem by analyzing what he considers some of the most crucial parameters in the process of conception: space and time, motion and location, causation and force interaction, and attention and viewpoint. He showed that the understanding of semantic structures involves the combination of these domains into an integrated whole. Our classification of possible situations agrees with this selection of structuring domains, and implements this idea from a general cognitive semantics perspective into a concrete CVS for human behavior interpretation. In our approach, knowledge representation is organized in a linear fashion, which ranges from objective knowledge extracted using vision processes (low-level), to uncertain, subjective knowledge based on attentional factors and behavioral models (high-level). It is structured as follows, see Table 3:

- The *Status* class contains semantic knowledge expressed by metric-temporal facts. Its three subclasses depend on the information provided by three considered trackers: body, agent, and face. The set of detectable occurrences defines the possible results for classification processes taking into account the spatial configuration at given time-steps (*Action*, *Expression*), and the analysis of the trajectories for the agents within a location using fuzzy models of human motion (*Activity*). This first set of facts is only applicable to pedestrians, the second one applies to general agents.
- The *ContextualizedEvent* class contains semantic knowledge at a higher level, now considering an occurrence as the interaction among semantic entities. This knowledge emerges after a process of contextualization of different sources of information, and facilitates both the anticipation of events and the reasoning of causation by analyzing situational models.



**Table 3.** Central part of the ontology which shows the taxonomy for the classification of Situations, see Section 4.2 for details.

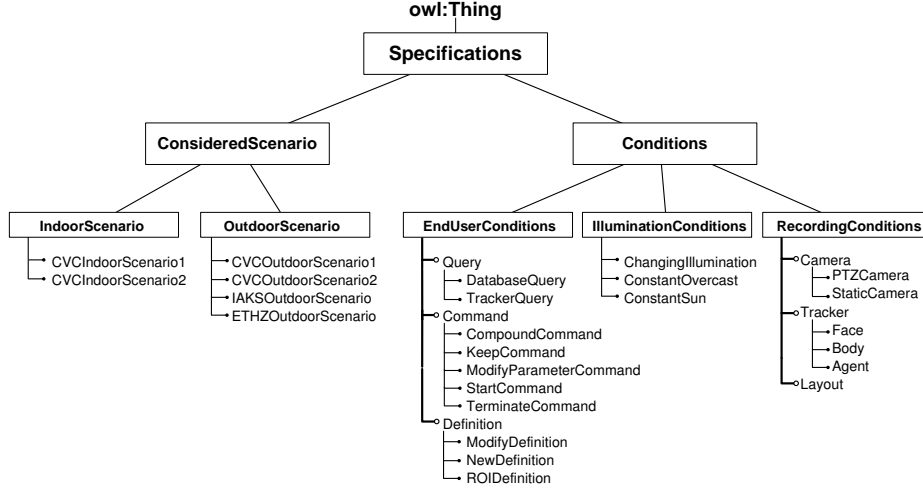
- Finally, the *BehaviorInterpretation* class specifies the set of interpretations which hold the greater level of uncertainty, and also claims for a bigger number of assumptions. Individuals in this class are selected regarding the intentional goals and attentional factors of the specific domain, in this case the detection of remarkable behaviors in urban outdoor scenarios involving humans and vehicular traffic, for surveillance purposes.

Each one of the described behaviors have been conditioned to the presence of certain arguments, characterized by the entities appearing in Table 2. For instance, a *DangerOfRunover* situation involves at least an Agent of type Vehicle and another Agent of type Pedestrian, and a *Theft* situation involves a minimum of two Pedestrians and a *PickableObject*.

A third taxonomy has been also incorporated into the ontology for practical considerations, see Table 4. These specifications are related to restraints for the particular scenarios chosen, regarding conditions such as illumination, recording, distributed camera layout, or possible kinds of commands and queries from a user given these characteristics. Each participant of the *ConsideredScenario* category is assigned to a set of *Conditions*. The incorporation of new scenarios into the HSE system is thus directly included within this part of the ontology.

## 5 Conclusions and Future Work

An ontology has been designed to account and organize the possible set of situations to be handled by a CVS based on HSE for surveillance applications. These situations are represented by means of HLSP, which hold the higher level



**Table 4.** Taxonomy for the organization of specifications related to the concrete implementation of the HSE system.

of semantics without being subjected to a specific language. As a result, the obtained ontology builds on a neutral framework which integrates semantic representations from vision and linguistics in a cognitive environment.

The modeling of this ontology permits to reduce the complexity associated to the multilingual dimension of the system. It also allows to clarify and simplify the design and implementation of components to bridge the semantic gap. The proposed ontology is particularly useful in the application field of NL understanding, since it makes easier the categorization of a discourse and plays a great role in disambiguation. Another important benefit is to restrict the domain of acceptance for the different forms of semantic representation, since the constraints applied to the taxonomy fix the validity of the situations to detect. This way, mechanisms for prediction based on restrained behavioral models can be developed.

One immediate application of the ontology is related to the field of semantic indexation. The ontology of situations provides the space and validity of possible annotations for video sequences related to the domain. In the specific implementation, a SGT acts as an actual content classifier, which characterizes the temporal structure of video sequences from a semantic perspective. Thus, the HLSP can be identified as high-level semantic indexes, which can facilitate further applications such as search engines and query-based retrieval of content.

Some important issues have to be covered in further steps. First, a proper communication with the NL interfaces requires to relate the proposed ontology of situations to a linguistic-oriented ontology. Initially, syntactical and grammatical models for each specific language to implement have to be defined, so that each

HLSP is representable by a set of LP in the target language. In addition to this, the domain of application for the surveillance system has to be enlarged, so that other outdoor and indoor situations in surveillance environments are possible. Thus, eventual extensions of the defined ontologies depend on the necessity to cover new situations.

## References

1. M. Arens and H.-H. Nagel. Behavioral Knowledge Representation for the Understanding and Creation of Video Sequences. volume 2821.
2. M. Arens, A. Ottlik, and H.-H. Nagel. Natural Language Texts for a Cognitive Vision System. *Proceedings of the 15th European Conference on Artificial Intelligence, ECAI'2002, Lyon, France, July 2002*, pages 455–459, 2002.
3. H. Buxton. Learning and Understanding Dynamic Scene Activity: a Review. *Image and Vision Computing*, 21(1):125–136, 2003.
4. B. Chandrasekaran, J.R. Josephson, and V.R. Benjamins. What Are Ontologies, and Why Do We Need Them? *Intelligent Systems and Their Applications, IEEE [see also IEEE Intelligent Systems]*, 14(1):20–26, 1999.
5. W. Croft and A.D. Cruse. *Cognitive Linguistics (Cambridge Textbooks in Linguistics)*. Cambridge University Press, January 2004.
6. J. L. Crowley. Situated observation of human activity. In C. Jaynes and R. Collins, editors, *1st Workshop on Computer Vision for Interactive and Intelligent Environment (CVIIIE 2005)*, pages 97–108, Lexington, Kentucky, USA, 17–18 November 2005. IEEE Computer Society, Los Alamitos, CA, USA.
7. R. Cucchiara. Multimedia Surveillance Systems. *Proceedings of the 3rd ACM International Workshop on Video Surveillance & Sensor Networks*, pages 3–10, 2005.
8. C. Fernández, P. Baiget, X. Roca, and J. González. Natural Language Descriptions of Human Behavior from Video Sequences. In *30th Annual German Conference on Artificial Intelligence (KI 2007)*. Springer LNCS, 2007.
9. J. González. *Human Sequence Evaluation: The Key-Frame Approach*. PhD thesis, Universitat Autònoma de Barcelona, Barcelona, Spain, 2004.
10. G. Granlund. *Cognitive Vision Systems*, chapter Organization of Architectures for Cognitive Vision Systems, pages 37–55. Springer Verlag, Heidelberg, 2006. Hans Hellmut Nagel and Henrik I. Christensen eds.
11. T.R. Gruber. Toward Principles for the Design of Ontologies Used for Knowledge Sharing? *International Journal of Human Computer Studies*, 43(5-6):907–928, 1995.
12. H. Kamp and U. Reyle. *From Discourse to Logic*. Kluwer Academic Publishers, Dordrecht Boston London, 1993.
13. A. Kojima and T. Tamura. Natural Language Description of Human Activities from Video Images Based on Concept Hierarchy of Actions. *International Journal of Computer Vision*, 50(2):171–184, 2002.
14. R. Langacker. Context, Cognition, and Semantics: A Unified Dynamic Approach. *Job*, 28:179–230, 2003.
15. M. Ma and P. Mc Kevitt. Visual semantics and ontology of eventive verbs. *Proceedings of the First International Joint Conference on Natural Language Processing*, pages 278–285, 2004.

16. N. Maillot, M. Thonnat, and A. Boucher. Towards Ontology Based Cognitive Vision. *Machine Vision and Applications*, 16(1):33–40, 2004.
17. S. Park and J.K. Aggarwal. Event Semantics in Two-person Interactions. *Proceedings of the Pattern Recognition, 17th International Conference on (ICPR'04)*, 4:227–230, 2004.
18. D. Rowe, I. Huerta, J. González, and J.J. Villanueva. Robust Multiple-People Tracking Using Colour-Based Particle Filters. In *3rd Iberian Conference on Pattern Recognition and Image Analysis (Ibpria 2007)*. Springer LNCS, 2007.
19. K. Schäfer and C. Brzoska. F-Limette Fuzzy Logic Programming Integrating Metric Temporal Extensions. *Journal of Symbolic Computation*, 22(5-6):725–727, 1996.
20. S. Staab and R. Studer. *Handbook on Ontologies*. Springer, 2004.
21. L. Talmy. *Toward a Cognitive Semantics - Volume 1: Concept Structuring Systems*. Bradford Book, 2000.
22. M. Vincze, W. Ponweiser, and M. Zillich. Contextual Coordination in a Cognitive Vision System for Symbolic Activity Interpretation. *Proceedings of the Fourth IEEE International Conference on Computer Vision Systems*, pages 12–12, 2006.